

پیش بینی مقاومت مارشال آسفالت با استفاده از الگوریتم های یادگیری ماشین نظارت شده ماشین بردار پشتیبان و جنگل تصادفی

مقاله علمی - پژوهشی

حسن حسین زاده*، دانش آموخته کارشناسی ارشد، مهندسی عمران، موسسه آموزش عالی اقبال لاهوری، مشهد، ایران
علیرضا حسینی، دانشجوی کارشناسی ارشد، مهندسی عمران، موسسه آموزش عالی اقبال لاهوری، مشهد، ایران
سینا آرمان، دانش آموخته کارشناسی ارشد، مهندسی عمران، دانشگاه آزاد اسلامی، مشهد، ایران
امیر سیاوش حجازی، دانش آموخته کارشناسی ارشد، مهندسی عمران، دانشگاه آزاد اسلامی، مشهد، ایران
*پست الکترونیکی نویسنده مسئول: Hosseinzadeh-Ha@mashhad.ir

دریافت: ۱۴۰۲/۰۱/۲۰ - پذیرش: ۱۴۰۲/۰۴/۲۸

صفحه ۲۶۲-۲۴۹

چکیده

سازمان‌های مسئول ساخت و نگهداری راه‌ها معمولاً از برخی معیارها برای واجد شرایط بودن مخلوط‌های آسفالتی قبل از استفاده در ساخت و ساز استفاده می‌کنند. یکی از مهم‌ترین ویژگی‌هایی که در طرح اختلاط و کنترل کیفی آسفالت سنجیده می‌شود مقاومت مارشال آسفالت می‌باشد. این مطالعه استفاده از روش‌های یادگیری ماشین را برای پیش بینی مقاومت مارشال آسفالت را بررسی می‌کند. با توجه به زمان‌بر بودن و هزینه‌بر بودن فرایند تولید و کنترل کیفی آسفالت، استفاده از روش‌های نوین در این فرایند ضرورت دارد. در این پژوهش از دو الگوریتم نظارت شده ماشین بردار پشتیبان و جنگل تصادفی که از الگوریتم‌های یادگیری ماشین محسوب می‌شوند به منظور پیش‌بینی مقاومت مارشال آسفالت استفاده شد. برای این منظور، نتایج آزمایشات ۲۰۰۰ نمونه آسفالت کارخانه آسفالت سازمان عمران شهرداری مشهد شامل دانه‌بندی مصالح، درصد شکستگی مصالح، درصد جذب قیر، وزن مخصوص قیر، وزن مخصوص حقیقی مصالح، درصد قیر مصرفی، نسبت درصد وزنی فیلر به قیر مؤثر و مقاومت مارشال آسفالت برای آموزش و ارزیابی مدل‌ها بکاررفته است. پس ساخت مدل و ارزیابی آن‌ها، مقدار R^2 برای روش ماشین بردار پشتیبان برابر ۸۷/۵ و برای جنگل تصادفی ۸۲/۶۹ به دست آمده است. همچنین مقادیر MAPE، RMSE و SDE برای SVM به ترتیب معادل ۳/۱۰۷۳، ۴۲/۰۴۲ و ۰/۰۲۰۸ و برای RF به ترتیب معادل ۳/۱۶۴۱، ۱/۸۷۰ و ۰/۰۲۱۱ محاسبه گشت. نتایج حاصل نشان دهنده کارآمدی مدل‌های استفاده‌شده در برابر روش‌های آزمایشگاهی برای پیش‌بینی مقاومت مارشال آسفالت است که روش SVM عملکرد مطلوب‌تری را نسبت به RF داراست. از روش‌های یادگیری ماشین می‌توان برای پیش‌بینی سایر پارامترهای طرح اختلاط آسفالت استفاده و زمان، هزینه و خطای انسانی آزمایشات را کاهش داد.

واژه‌های کلیدی: آسفالت، الگوریتم‌های یادگیری ماشین، جنگل تصادفی، ماشین بردار پشتیبان، مقاومت مارشال

۱- مقدمه

از جمله خستگی و ترک‌ها می‌شود. به‌طور کلی می‌توان گفت که مقدار مقاومت آسفالت نشان‌دهنده توانایی مخلوط آسفالتی برای مقاومت در برابر شیار شدگی و برآمدگی تحت بارگذاری سنگین ترافیکی است (Ali, 2006; Choubane et al.,

از مهم‌ترین ویژگی‌هایی که در اختلاط آسفالت مورد توجه قرار می‌گیرد مقاومت مارشال آسفالت است (Ozgan, 2009; Ozgan & Saruhan, 2010). پایین بودن مقدار مقاومت مارشال آسفالت باعث پایین آمدن کارایی آن و ایجاد مشکلاتی

تجربه E را شکل می‌دهد، اگر عملکرد آن در گروه وظایف T آن‌طور که توسط P اندازه‌گیری شده باتجربه E بهبود پیدا کند (Mitchell, 1997). به عبارت ساده‌تر، اگر یک برنامه کامپیوتری بتواند عملکرد خود در انجام یک وظیفه را با استفاده از تجربیات قبلی‌اش بهبود ببخشد آنگاه می‌توانید بگویید که آن ماشین یاد گرفته است؛ بنابراین، یادگیری ماشین عبارت از بهینه سازی یک عملکرد با استفاده از تجربیات گذشته است. فرآیند آموزش معمولاً با به‌کارگیری اطلاعات تجربی از هدف موردبررسی صورت گرفته و پس از ایجاد مدل آموزشی، از آن به‌عنوان مدلی جهت پیش‌بینی و یا طبقه‌بندی رفتار سیستم موردبررسی استفاده می‌شود (Burkov, 2020; Kanevski, 2009; Zhang, 2020). الگوریتم‌های یادگیری ماشین بر اساس نوع نحوه آموزش به چهار نوع نظارت‌شده، نظارت‌نشده، نیمه نظارتیو یادگیری تقویت شونده تقسیم می‌شوند. در الگوریتم‌های نظارت‌شده با دو نوع از متغیرها سروکار داریم. نوع اول که متغیرهای مستقل نامیده می‌شوند، یک یا چند متغیر هستند که قرار است بر اساس مقادیر آن‌ها، یک متغیر دیگر را پیش‌بینی کنیم. نوع دوم هم متغیرهای وابسته یا هدف یا خروجی هستند و قرار است مقادیر آن‌ها را به کمک این الگوریتم‌ها پیش‌بینی کنیم. برای این منظور باید تابعی ایجاد شود که ورودی‌ها (متغیرهای مستقل) را گرفته و خروجی موردنظر (متغیر وابسته یا هدف) را تولید کند. فرآیند یافتن این تابع که در حقیقت کشف رابطه‌ای بین متغیرهای مستقل و متغیرهای وابسته است، فرآیند آموزش نامیده می‌شود که روی داده‌های موجود (داده‌هایی که هم متغیرهای مستقل و هم متغیرهای وابسته آن‌ها معلوم است) اعمال می‌شود و تا رسیدن به دقت لازم، ادامه می‌یابد (Chourasiya & Jain, 2019; Goodfellow et al., 2016; Zhang, 2020).

الگوریتم‌های درخت‌های تصمیم، جنگل‌های تصادفی، K نزدیک‌ترین همسایه ماشین بردار پشتیبان، روش‌های مبتنی بر نظریه بیز و رگرسیون لجستیک نمونه‌ای از این یادگیری می‌باشند (Jordan & Mitchell, 2015). مطالعات متعددی در خصوص کاربرد روش‌های هوشمند در جهت پیش‌بینی پارامترهای مختلف آسفالت انجام شده است. ازگان در سال ۲۰۰۹، به‌منظور پیش‌بینی مقاومت مارشال آسفالت از روش منطق فازی استفاده کرد. در این مطالعه از نتایج ۶۵ نمونه آسفالت استفاده شد و نتایج نشان داده شد که روش منطق

بتن آسفالتی از یک استخوان‌بندی مصالح سنگی خوب دانه‌بندی شده با فضای خالی مناسب که قیر سطح دانه‌ها را اندود کرده و آن‌ها را به یکدیگر چسبانده است تشکیل شده و مرغوب‌ترین نوع آسفالت به شمار می‌رود. مقاومت مارشال بتن آسفالتی نیز تابعی از دانه‌بندی، ویژگی‌های مصالح سنگی و همچنین درصد و نوع قیر مصرفی آن است (Özgan & Saruhan, 2010). به دلیل متغیرها و پیچیدگی روابط بین آن‌ها و مقاومت مارشال، تاکنون رابطه تحلیلی ریاضی بین این متغیرها و مقاومت مارشال و تأثیر متقابل آن‌ها ارایه نگردیده و صرفاً به کمک آزمایش تعیین می‌گردد. از آنجایی که ساخت و توسعه راه‌ها از مبانی پیشرفت اقتصادی و فرهنگی کشورها محسوب می‌شود و با توجه به هزینه‌های سنگین آسفالت و نگهداری آن ضرورت استفاده از روش‌های جدید و پیشرفته‌تر در طرح و کنترل کیفی آسفالت روزبه‌روز محسوس‌تر می‌شود (Ozgan, 2009; Özgan & Saruhan, 2010). حجم بالای داده‌ها در مطالعات و ارتباط پیچیده موجود بین معیار باعث شده تا روش‌های تحلیل سنتی دارای عدم قطعیت بالایی باشند. از این رو به‌کارگیری روش‌هایی که توانایی تحلیل و کشف ارتباط پیچیده فضایی را در بین داده‌های اولیه دارا باشند توسعه یافته است. از جمله این روش‌ها می‌توان به خانواده روش‌های مبتنی بر یادگیری ماشین و تشخیص الگو اشاره کرد که به‌کارگیری آن‌ها بخصوص در سالیان اخیر محبوبیت بسیاری در بین دانشمندان پیدا کرده است. در دو دهه گذشته الگوریتم‌های هوش مصنوعی در حل مسائل مرتبط با علوم مهندسی به‌صورت وسیعی مورد استفاده قرار گرفته است (Lary et al., 2016; Majidifard et al., 2020). الگوریتم‌های آموزش ماشین به‌عنوان زیرشاخه‌ای از علم هوش مصنوعی، به‌منظور دریافت جوابی با درصد اطمینان قابل قبول توسعه یافته است. هدف از آموزش ماشین در واقع بهینه‌سازی عملکرد سیستم با استفاده از تجربیات گذشته است (Kanevski, 2009; Wang et al., 2009). یادگیری ماشین یکی از مهم‌ترین بخش‌های هوش مصنوعی است که از علوم کامپیوتر آغاز شده و به‌سرعت در حال رشد است به‌طوری‌که در حوزه‌های دیگر علوم نیز کاربردهای فراوانی یافته است. یکی از تعاریف یادگیری ماشین آن‌طور که از سوی تام میشل ارایه گردیده است بدین شرح است: نوعی برنامه کامپیوتری که با توجه به برخی وظایف گروه T و عملکرد P،

حالت با استفاده از توابع کرنل، داده‌ها به فضایی به بعد بزرگ‌تر انتقال می‌یابند که در آن تفکیک‌پذیری بهتر انجام می‌شود و ابر صفحه جداکننده در آن فضا تعیین می‌شود. در روش ماشین بردار پشتیبان از چهار تابع کرنل خطی، چندجمله‌ای، تابع پایه شعاعی و تابع سیگموئید بیشتر استفاده می‌شود. روابط این توابع در جدول ۱ آورده شده است.

جدول ۱. توابع کرنل متداول در ماشین بردار پشتیبان (Shawe-Taylor & Cristianini, 2000)

نوع تابع	فرمول
خطی	$x_i \times x_j$
چندجمله‌ای	$[\gamma x_i x_j + \text{coefficient}]^d$
پایه شعاعی	$\exp(-\gamma \ x_i - x_j\ ^2)$
سیگموئید	$\tanh(\gamma x_i x_j + \text{coefficient})$

الگوریتم‌های مبتنی بر ماشین‌های بردار پشتیبان الگوریتم‌هایی هستند که سعی می‌کنند یک حاشیه را بیشینه کنند. این الگوریتم‌ها برای پیدا کردن خط جداکننده دسته‌ها، از دو خط موازی شروع کرده و این خطوط را در جهت خلاف یکدیگر حرکت می‌دهند تا هرکدام از خطوط به یک نمونه از یک دسته خاص در سمت خود برسد. پس از انجام این مرحله، میان دو خط موازی یک نوار یا حاشیه شکل می‌گیرد. هرچه پهنای این نوار بیشتر باشد، به این معناست که الگوریتم توانسته حاشیه را بیشینه کند و هدف نیز بیشینه نمودن این حاشیه است. در واقع هدف ما این است که بیشترین مقدار ممکن برای این حاشیه انتخاب شود. هرچه حاشیه بیشتر باشد خطای کلاس‌بندی کمتر خواهد بود. در مرکز حاشیه شکل، خط جداکننده دسته‌ها یا همان خط مرکزی (در حالت دوعبدهی خط، سه‌بعدی صفحه و در حالت بیشتر از سه‌بعدی ابر صفحه) قرار می‌گیرد. حال از بین خطوطی که رسم می‌شوند، الگوریتم خطی را که حاشیه کناری آن بیشتر باشد، به‌عنوان خط جداکننده دسته‌ها انتخاب می‌کند. حاشیه مربوط به دو خط B_1 و B_2 در شکل ۱ نمایش داده شده است. الگوریتم خط B_1 را به‌عنوان خط جداکننده انتخاب می‌کند چراکه حاشیه کناری این خط، نسبت به حاشیه کناری خط B_2 بیشتر است.

فازی و روش‌های آماری می‌توانند مقاومت مارشال آسفالت را با دقت خوبی پیش‌بینی کند (Ozgan, 2009). در مطالعه‌ای در سال ۲۰۱۶ از شبکه عصبی مصنوعی جهت پیش‌بینی مقاومت مارشال آسفالت حاوی پلی پارافینلین استفاده شد. در این تحقیق که توسط کاراهانسر و همکاران انجام شد، مدل با توجه به درصد پلی پارافینلین، درصد قیر، روانی و حجم نمونه آموزش داده شد و سپس مقاومت مارشال پیش‌بینی شد که میزان صحت مدل در این مطالعه ۶۹ درصد به دست آمد (Karahancer et al., 2016). عمر و همکاران در سال ۲۰۱۸، از مدل رگرسیون چندگانه اقدام به پیش‌بینی مقاومت مارشال آسفالت کردند. مدل‌های مجزایی بر اساس یک متغیر تا شش متغیر ساخته شد که دقت پیش‌بینی آن‌ها ۰.۵۱ تا ۰.۵۷۱ متغیر بود (Cansiz & Askar, 2018).

نکته مشترک تمامی این مطالعات، دقت پایین نتایج به‌دست‌آمده است و بدین معناست که این روش‌ها نتوانسته‌اند به‌خوبی رابطه پیچیده بین متغیرها را تشخیص دهند. در این تحقیق از الگوریتم رگرسیون ماشین بردار پشتیبان و جنگل تصادفی به‌منظور پیش‌بینی مقاومت مارشال آسفالت استفاده شده است. این دو الگوریتم در علوم مهندسی و سایر علوم بسیار پرکاربرد بوده‌اند و عملکرد مثبتی داشته‌اند (Aroef et al., 2020; Cardoso-Fernandes et al., 2019). این مطالعه به‌منظور پیش‌بینی مقاومت آسفالت استفاده شده است. به‌منظور دستیابی به این هدف از داده‌های آزمایشگاه کنترل کیفیت سازمان عمران شهرداری مشهد استفاده شده است که در ادامه توضیح داده می‌شود.

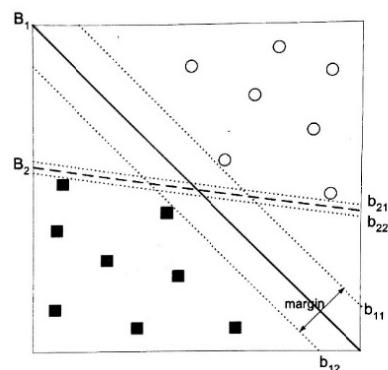
۲- مواد و روش‌ها

۲-۱- ماشین بردار پشتیبان

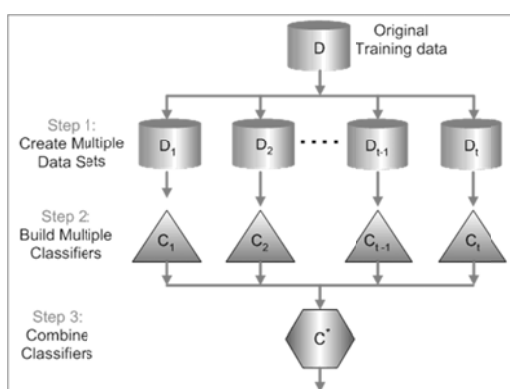
استفاده از بردارهای پشتیبان در مسائل دسته‌بندی، رویکرد جدیدی است که در چند سال اخیر موردتوجه بسیاری قرار گرفته است. ماشین بردار پشتیبان در ابتدا توسط واپنیک در سال ۱۹۹۰ طراحی شد و نظریه آماری یادگیری را به‌صورت مستحکم بنا نهاد. در تعریف ساده، SVM الگوریتمی است که به‌وسیله یک ابرصفحه جداکننده که روی داده‌های آموزشی تعریف می‌شود، دو کلاس خاص از هم جدا و مشخص می‌شوند (Cortes & Vapnik, 1995; Yang et al., 2008). از جمله قابلیت‌های ماشین بردار پشتیبان، توانایی غلبه بر مشکل توزیع غیرخطی داده‌های آموزشی است. در این

شده است. پارامترهای مهم در الگوریتم جنگل تصادفی، تعداد متغیر انتخاب شده در هر گره درخت و تعداد درخت می‌باشند که لازم است برای آن‌ها مقادیر مناسبی تعیین شوند (Harris & Grunsky, 2015; Oshiro et al., 2012). در این روش در هر بار نمونه‌گیری با جایگزینی اطلاعات، برخی اطلاعات هرگز نمونه‌گیری نمی‌شوند و برخی دیگر شاید چند بار نمونه‌گیری شوند.

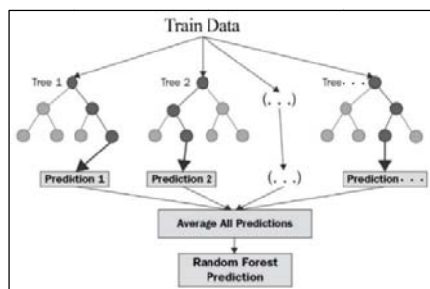
به عبارت دیگر برخی داده‌های ورودی برای برخی درخت‌ها در اصطلاح نمونه خارج از کیسه (OOB) خواهند بود یعنی در ایجاد برخی درخت‌ها مشارکت نخواهند داشت.



شکل ۱. حاشیه خطوط دسته‌بندی نمونه



شکل ۲. ایده کلی الگوریتم‌های تلفیقی



شکل ۳. ایده الگوریتم جنگل تصادفی

این داده‌ها عمل یک اعتبارسنج داخلی برای هر درخت را دارند که این اعتبارسنجی از طریق برآورد خطای خارج از کیسه انجام می‌شود. اگر خود داده‌های خارج از کیسه از طریق درختان پیش‌بینی شوند، برای این پیش‌بینی‌ها خطا وجود خواهد داشت و میانگین این خطاها، خطای خارج از کیسه نامیده می‌شود که نشان‌دهنده میزان تأثیر نمونه‌های انتخاب‌نشده بر میزان خطای نتیجه نهایی جنگل تصادفی است (Gareth et al., 2013).

برای به دست آوردن نتایج مناسب کلاسه‌بندی در روش ماشین بردار پشتیبان این نکات را بایستی موردتوجه قرارداد: به دست آوردن مقدار بهینه برای پارامتر C و تغییر پارامترهای مختلف تابع کرنل یا تعویض آن‌ها با استفاده از روش‌های Cross Validation (Jakkula, 2006; Shabankareh & Hezarkhani, 2017).

۲-۲- جنگل تصادفی

روش جنگل تصادفی از جمله روش‌های تلفیقی است. در روش‌های تلفیقی مدلی که برای دسته‌بندی یا رگرسیون انتخاب می‌شود که ترکیبی از چندین مدل است. هر مدل رأی خود را صادر می‌کند و نتیجه نهایی در مورد مقدار بر اساس این رأی‌ها صادر می‌شود. اغلب صحت یک روش تلفیقی بهتر از مؤلفه‌های تشکیل‌دهنده آن یعنی مدل‌ها است (Mohri et al., 2018). ایده کلی یک الگوریتم تلفیقی در شکل ۲ آورده شده است. تصور کنید مدل‌های استفاده‌شده در روش تلفیقی همگی از نوع درخت تصمیم هستند، بدین ترتیب این مجموعه تشکیل یک جنگل را خواهد داد. هر یک از درختان تصمیم با استفاده از یک انتخاب تصادفی صفات خاصی موجود در هر گره جهت تعیین انشعاب ساخته می‌شوند.

به عبارت دیگر هر درخت بر اساس مقادیر یک بردار تصادفی ساخته می‌شود. این مقادیر دارای توزیع یکسانی برای تمام درختان موجود در جنگل هستند و به صورت مستقلی نمونه‌گیری می‌شوند. برای دسته‌بندی نیز هر درخت رأی خود را صادر و نتیجه نهایی برآی اکثریت تعیین می‌شود (Liaw & Wiener, 2002). ایده جنگل تصادفی در شکل ۳ نشان داده

۳- ارزیابی مدل

ارزیابی عملکرد مدل‌های SVM و RF بر اساس ضریب تعیین، میانگین مربعات خطای تخمین، میانگین خطای نسبی و قدر مطلق خطای نسبی که همگی معرف میزان دقت (خطای) تخمین می‌باشند استفاده شده است. مدلی که بیشترین مقدار R^2 و کمترین مقدار RMSE، SDE و MAPE را دارا باشد از عملکرد مناسب‌تری برخوردار است. کمیت R^2 که ضریب تعیین نیز نامیده می‌شود، نسبت تغییرات (متغیرهای) تعریف‌شده، به کل تغییرات (متغیرها) می‌باشد. این اندازه‌گیری به ما این امکان را می‌دهد که تعیین کنیم چقدر می‌توان به پیش‌بینی یک مدل یا نمودار مطمئن بود. روش‌های دیگر اگرچه مفید هستند، اما یک اشکال مشترک دارند: از آنجاکه مقادیر آن‌ها می‌تواند بین صفر تا بی‌نهایت متغیر باشد، یک مقدار واحد آن‌ها در مورد عملکرد به‌تنهایی چیز زیادی نمی‌گوید و فقط در هنگام مقایسه قابل تفسیر هستند (Chicco et al., 2021).

در این پژوهش چون نوع پیش‌بینی رگرسیونی است، برای رد یا قبول نتایج می‌بایست از معیار R^2 استفاده کرد و برای مقایسه مدل‌های استفاده‌شده با یکدیگر می‌توان از سایر روش‌های ذکرشده استفاده کرد.

$$MAPE = \frac{100}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i} \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (7)$$

$$SDE = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{|y_i - \hat{y}_i|}{y_i} - \frac{MAPE}{100} \right)^2} \quad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (9)$$

y_i ، \hat{y}_i و \bar{y}_i به ترتیب خروجی مطلوب، خروجی تخمینی توسط مدل و میانگین خروجی‌های مطلوب می‌باشند و N نیز تعداد داده‌ها می‌باشد.

۴- جمع‌آوری، آماده‌سازی و پردازش داده مورد نیاز

در این مقاله برای بررسی عملکرد الگوریتم‌های SVM و RF در تخمین مقاومت فشاری آسفالت از نتایج ۲۰۰۰ نمونه آزمایشگاهی بتن آسفالتی آزمایشگاه کنترل کیفیت سازمان عمران شهرداری مشهد استفاده گردیده است. این نتایج مربوط به نمونه‌های اخذشده از آسفالت‌های تولیدشده در کارخانه آسفالت شهید مجیدی سازمان عمران در سال‌های ۱۳۹۹ و ۱۴۰۰ است. داده‌های منتج از نمونه‌ها به‌صورت برخط در سامانه جامع مدیریت داده ثبت می‌شود. نتایج آزمایشگاهی شامل دانه‌بندی مصالح (ASTM D5444-15, 2015)، درصد شکستگی (ASTM D5821-13, 2017)، درصد جذب قیر (ASTM D4469-17, 2017)، وزن مخصوص قیر (ASTM D70, 2021)، وزن مخصوص حقیقی مصالح (ASTM C29, 2017)، درصد قیر (ASTM D2172, 2017) نسبت فیلر به قیر مؤثر و مقاومت مارشال (ASTM D6927-15, 2015) است. فضای خالی که نقش مهمی در عملکرد مقاومت روسازی دارد، به دلیل اینکه مقدار آن برای بیش از ۳۰ درصد از نمونه‌ها ثبت نشده بود و در صورت استفاده از روش‌های جایگذاری داده سنسورد یک داده مصنوعی تولید می‌شد از مجموع داده‌ها حذف گردید.

دانه‌بندی یک مخلوط مصالح سنگی عموماً دارای سه بخش اصلی است: درشت‌دانه‌ها، ریزدانه‌ها و پرکننده‌های معدنی. طبق استاندارد ASTM D3515 و آسفالت‌های مورد استفاده، درصد عبوری الک‌های (اینچ، ۳/۴ اینچ، ۱/۲ اینچ، ۳/۸ اینچ، شماره ۴/۷۵ میلی‌متر)، شماره ۸ (۲/۳۶ میلی‌متر)، شماره ۵۰ (۰/۳ میلی‌متر) و شماره ۲۰۰ (۰/۰۷۵ میلی‌متر) در آزمایش اندازه‌گیری می‌شوند. دانه‌بندی نمونه‌های استفاده‌شده در این تحلیل منطبق با رده D-4 و D-5 جدول یک استاندارد ASTM D3515 می‌باشند که D-4 عمدتاً برای اساس قیری و آستر و D-5 برای عمدتاً آستر و رویه استفاده می‌شوند. حدود استاندارد ذکرشده و مشخصات دانه‌بندی کارگاهی طرح اختلاط مربوط به نمونه‌ها در جدول ۲ آورده شده است.

قیر استفاده از شده در تمامی نمونه‌ها از نوع قیر درجه عملکردی PG64-22 است که طی آزمایش‌های ادواری کنترل کیفیت، نتایج همیشه مطابق با مشخصات مندرج در استاندارد ASTM D6373 بوده است. مصالح سنگی مورد استفاده در

خارج از ردیف توسط مکانیسم متفاوت نسبت به سایر داده‌ها خواهد بود. خطاهای انسانی در اندازه‌گیری، آماده‌سازی، آنالیز و ثبت یکی از مهم‌ترین این مکانیسم می‌تواند باشد (Aggarwal, 2017; Hawkins, 1980).

هرگونه پردازش و تفسیر داده‌ها بدون در نظر گرفتن این مسئله می‌تواند باعث ایجاد خطا و اشتباه در نتایج گردد. بنابراین شناسایی داده‌های خارج از ردیف اولین گام در پردازش داده‌ها محسوب می‌شود.

نمونه‌ها از جنس آهنی بوده و از یک معدن در شمال مشهد تهیه می‌شود که مشخصات شیمیایی و فیزیکی آن‌ها منطبق با استانداردهای درج‌شده در مراجع است.

پس از جمع‌آوری نتایج، می‌بایست پیش‌پردازش بر روی داده انجام شود و داده‌های پرت از مجموعه داده‌ها حذف گردد. داده خارج از ردیف (داده پرت) به داده‌ای اطلاق می‌شود که به‌طور آشکار از سایر داده‌ها فاصله‌ی معنی‌داری داشته باشد؛ بنابراین فاصله زیاد نشان‌دهنده‌ی تشکیل داده‌ی

جدول ۲. حدود دانه‌بندی استاندارد و کارگاهی مرتبط با طرح اختلاط نمونه‌ها

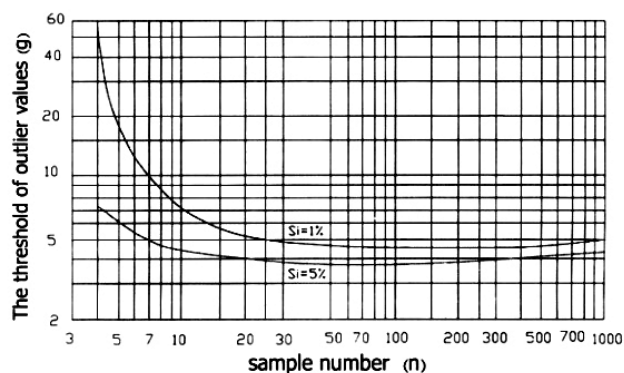
	۲۵	۱۹	۱۲/۵	۹/۵	۴/۷۵	۲/۳۶	۳۰۰	۷۵
	میلی‌متر (1 in)	میلی‌متر (3/4 in)	میلی‌متر (1/2 in)	میلی‌متر (3/8 in)	میلی‌متر (No.4)	میلی‌متر (No.8)	میکرومتر (No.50)	میکرومتر (No.200)
D-4	۱۰۰	۹۰-۱۰۰	---	۵۶-۸۰	۳۵-۶۵	۲۳-۴۹	۵-۱۹	۲-۵
D-5	---	۱۰۰	۹۰-۱۰۰	---	۴۴-۷۴	۲۸-۵۸	۵-۲۱	۲-۱۰
توپکا-۰۱۹	---	۱۰۰	۹۴	---	۵۸	۳۶	۱۱	۴
بیندر-۰۱۹	۱۰۰	۱۰۰	---	۷۰	۴۹	۳۵	۱۲	۴/۵
بیندر-۰۲۵	۱۰۰	۹۴	---	۷۰	۴۹	۳۵	۱۱	۴/۵

برای اعمال این روش، میانگین (\bar{x}) و انحراف معیار داده‌ها (S) بدون در نظر گرفتن بزرگ‌ترین مقدار داده‌ها محاسبه می‌شوند. سپس بزرگ‌ترین مقدار داده‌ها (x_A) در صورتی در رابطه صدق کند یک مقدار خارج از ردیف در نظر گرفته می‌شود که در آن g حد آستانه‌ای مقادیر خارج از ردیف است که از نمودار شکل ۴ به دست می‌آید (Liu, 1998). این روش با کد نویسی در نرم‌افزار R نسخه ۱-۲-۴ بر روی داده‌ها اجرا گردیده است. لازم به ذکر است در مورد داده‌های پرتی که در شیت‌های آزمایشگاهی دارای نظر کارشناسی قابل استناد در توجیه پرت بودن داده داشتن تغییری صورت نگرفت.

به‌طورکلی روش‌های شناسایی داده‌های خارج از ردیف را می‌توان به دو گروه ۱- روش‌های بصری بر اساس نمودار (مثل روش نمودار احتمال، نمودار جعبه‌ای و نمودار خی‌دو) و ۲- روش‌های محاسباتی بر اساس تعیین آستانه (مثل روش دورفل، روش فانوپ و روش فرکتالی) تقسیم‌بندی نمود (Filzmoser et al., 2005; Wellmer, 2012; Zhang et al., 2015).

در این تحقیق مقادیر خارج از ردیف با استفاده از رابطه ۱۰ و نمودار دورفل در سطح اعتماد ۹۵٪ شناسایی و تبدیل شده‌اند.

$$x_A \geq \bar{x} + s \cdot g \quad (10)$$



شکل ۴. حد آستانه‌ای مقادیر خارج از ردیف (g) به‌عنوان تابعی از تعداد نمونه (n) و سطح اعتماد

پارامترهای C و γ متعلق به تابع کرنل مقدار C بین ۱ تا ۱۰۰ و مقدار γ بین ۰ تا ۱ در نظر گرفته شد و سپس از روش k -fold cross validation استفاده شده است. در این روش به ازای هر ترکیب C و γ در بازه‌های تعریف شده، داده‌ها به K زیرمجموعه افزایش می‌شوند. از این K زیرمجموعه، هر بار یکی برای اعتبارسنجی و $K-1$ تای دیگر برای آموزش بکار می‌روند. این روال K بار تکرار می‌شود و همه داده‌ها دقیقاً یکبار برای آموزش و یکبار برای اعتبارسنجی بکار می‌روند. در نهایت میانگین نتیجه این K بار اعتبارسنجی به عنوان یک تخمین نهایی برگزیده می‌شود. به طور معمول از روش اعتبارسنجی پنج لایه یا ده لایه در پژوهش‌های مدل‌سازی و پیش‌بینی استفاده می‌شود (Refaeilzadeh et al., 2009) که در این تحقیق از 10-fold استفاده شده است. براساس نتایج به دست آمده، در شکل ۵ مقدار بهینه C برابر ۹ و مقدار بهینه γ برابر ۰/۱۱ در نظر گرفته شده است. مقدار R^2 برای مدل SVM فقط با این ترکیب C و γ به مقدار بیشینه خود می‌رسد و مقادیر غیر از این دو باعث کاهش دقت و صحت الگوریتم می‌شود.

جدول ۳. جدول آمار توصیفی نمونه‌های مورد در تحلیل

شاخص	دامنه	مینیمم	ماکزیمم	میانگین
درصد عبوری از الک ۳/۴	۱۰	۹۰	۱۰۰	۹۷/۵
درصد عبوری از الک ۱/۲	۳۱/۵	۶۷/۵	۱۰۰	۸۷/۶۳
درصد عبوری از الک ۳/۸	۳۸/۴	۵۸/۸	۹۷/۲	۷۷/۸۵
درصد عبوری از الک ۴	۳۵/۷	۳۷/۸	۷۳/۵	۵۱/۱۲
درصد عبوری از الک ۸	۲۸/۲	۲۳	۵۱/۲	۳۳/۲
درصد عبوری از الک ۵۰	۱۴/۳	۵/۷	۲۰	۱۱/۱
درصد عبوری از الک ۲۰۰	۹/۹	۱/۱	۱۱	۶/۵۱
درصد شکستگی مصالح	۲۵	۷۵	۱۰۰	۸۹/۶
درصد جذب قیر مصالح	۰/۱۵	۰/۲۵	۰/۴	۰/۳۴
وزن مخصوص قیر	۰/۰۳	۱	۱/۰۳	۱/۰۱۵
وزن مخصوص حقیقی مصالح	۰/۳۳	۲/۴۹۵	۲/۸۲۵	۲/۷۶
درصد قیر مصرفی	۲/۳	۳/۴	۵/۷	۴/۵۱
فیلر به قیر مؤثر	۰/۴۷	۰/۸۵	۱/۳۲	۱/۰۳
مقاومت مارشال	۸۳۵	۶۹۵	۱۸۰۰	۱۰۰۰,۳۵

جدول آمار توصیفی نمونه‌ها پس از تعدیل داده‌های خارج از ردیف در جدول ۳ ارائه شده است. در زمینه‌ی الگوریتم‌های یادگیری ماشین هر نمونه، یک بردار ویژگی در نظر گرفته می‌شود که محتوای آن توسط نتایج آزمایش‌ها تعریف شده‌اند (Liu & Motoda, 2012)؛ بنابراین تعداد ابعاد بردار ویژگی برابر با تعداد آزمایش‌ها است. نتایج تمام آزمایش‌ها بعد از پردازش برای ایجاد ۲۰۰۰ بردار ویژگی ۱۳ بعدی با یکدیگر ترکیب شدند. داده‌ها بر اساس نتیجه مقاومت مارشال به دست آمده برچسب‌گذاری شدند. از این تعداد داده یک مجموعه شامل ۱۴۰۰ داده یعنی ۷۰ درصد کل مقادیر به صورت تصادفی به عنوان داده آموزشی و ۳۰ درصد باقی‌مانده به عنوان داده آزمایشی برای بررسی صحت تابع به دست آمده استفاده شده است. محققان نشان داده‌اند که این ترکیب بهترین عملکرد را به همراه داشته است (Ghorbanzadeh et al., 2018; Tiwari et al., 2018). با توجه به اینکه رنج داده‌ها و واحد اندازه‌گیری آن‌ها در لایه‌های اطلاعاتی متفاوت می‌باشند، در الگوریتم‌های یادگیری ماشین عملکرد تابع با محدودیت سخت‌افزاری و نرم‌افزاری همراه است و نمی‌تواند عملکرد مثبتی داشته باشد. برای مثال، در این پژوهش داده‌های وزن مخصوص قیر مقادیری بسیار کوچک‌تر از داده‌های دانه‌بندی دارند؛ بنابراین نیاز است تا مقادیر پردازش شده نرمالایز شوند (Jayalakshmi & Santhakumaran, 2011).

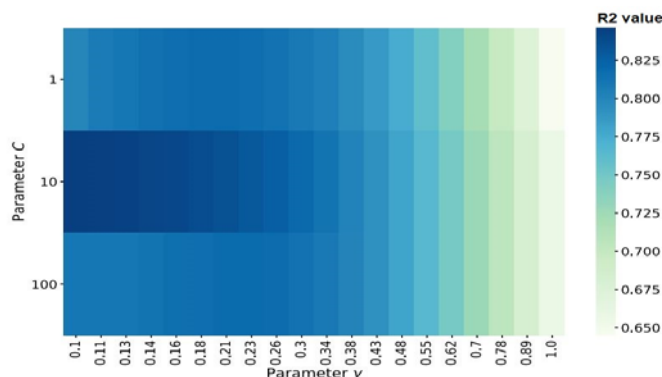
ساده‌ترین روش برای بی‌مقیاس کردن این است که مقادیر را به مقیاس $[0,1]$ تغییر دهیم. فرمول این روش به صورت زیر است:

$$\hat{x} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (11)$$

که در آن x داده اصلی و \hat{x} مقدار نرمالایز شده است (Mohamad & Usman, 2013).

۵- نتایج حاصل از مدل‌سازی با استفاده از روش SVM و RF

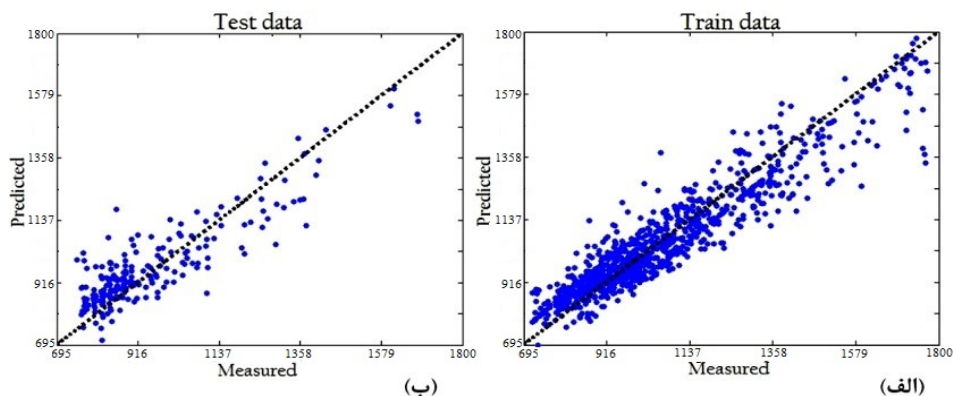
از داده‌های آموزشی برای ساخت مدل‌های SVM استفاده شده است. همان‌طور که در قبل اشاره شد در این روش می‌بایست از توابع کرنل استفاده کرد. مدل‌سازی بر روی داده‌های آموزشی توسط هر چهار تابع کرنل معرفی شده در جدول ۱ انجام شده است که بهترین نتایج با تابع کرنل پایه شعاعی به دست آمده است. همچنین به منظور بهینه کردن



شکل ۵. مقادیر R^2 برای ترکیب‌های مختلف C و γ با استفاده از روش 10-fold cross validation

درصد شود تمامی مقادیر مشاهده شده با مقادیر برازش شده یکسان خواهند بود و همه نقاط داده‌ها بر روی خط برازش شده قرار خواهند گرفت (Devore, 2011). کلیه فرایند مدل‌سازی با استفاده از بسته E1071 در نرم‌افزار R نسخه ۴-۲-۱ (Meyer et al., 2019) انجام شده است.

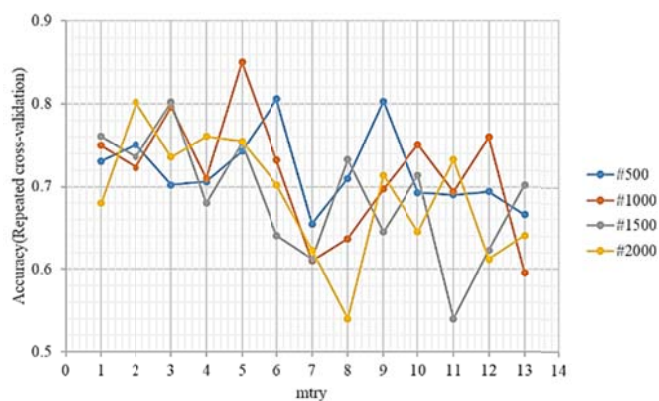
در شکل ۶ نمودار مقادیر واقعی اندازه‌گیری شده و مقادیر پیش‌بینی شده مقاومت فشاری آسفالت توسط مدل SVM برای داده‌های آموزشی و آزمایشی ترسیم شده است. همان‌طور که در شکل نیز به‌خوبی دیده می‌شود، نتایج SVM به خط نیم‌ساز نزدیک هستند و می‌توان گفت مقادیر را با دقت خوبی پیش‌بینی کرده است. از لحاظ تئوری اگر R^2 برابر ۱۰۰



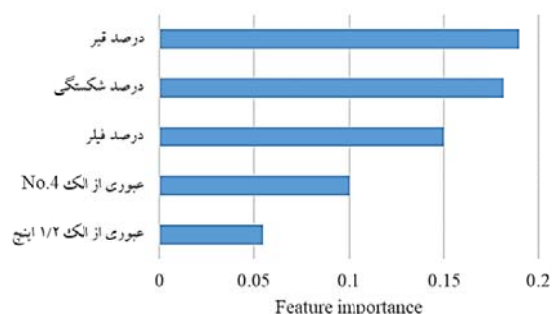
شکل ۶. منحنی پراکندگی مقاومت فشاری پیش‌بینی شده در مقابل مقاومت فشاری واقعی در مراحل آموزش (الف) و آزمایش (ب) در روش SVM

مسئله مشخص کند. در این تحقیق اهمیت پارامترهای ورودی بر اساس شاخص اهمیت جینی به دست آمده است (Han et al., 2016) که درصد قیر و شکستگی مصالح دارای اهمیت بالاتری نسبت به سایر پارامترها می‌باشند. در شکل ۸ اهمیت محاسبه شده برای پنج پارامتر دارای بیشترین اهمیت ارایه شده است.

در روش RF به‌منظور تعیین تعداد متغیر انتخاب شده در هر گره درخت (mtry) و همچنین تعداد درخت (ntree) از روش 10-fold cross validation استفاده شده است. مطابق شکل ۷ تعداد ۱۰۰۰ درخت و ۵ متغیر در هر گره مطلوب‌ترین شرایط به برای مدل فراهم کرده است. لذا این مقادیر به‌منظور مدل‌سازی استفاده شد. یکی از مزایای الگوریتم جنگل تصادفی این است می‌تواند اهمیت متغیرها را در یک



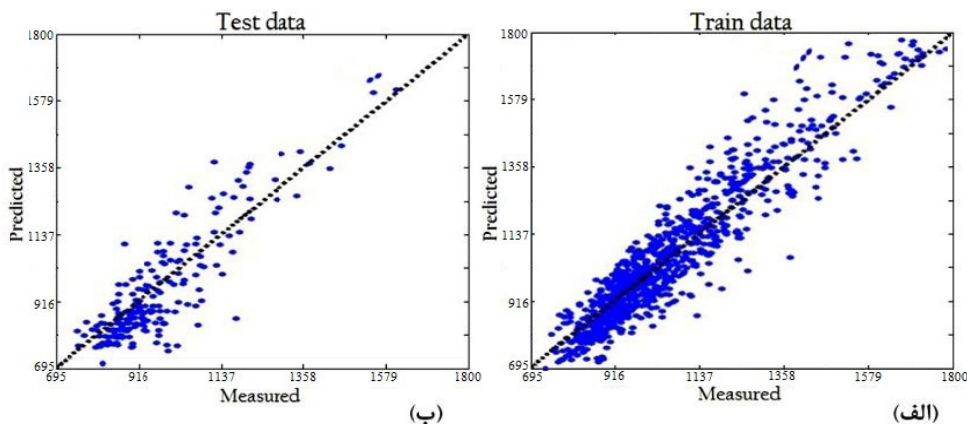
شکل ۷. نتایج cross validation به منظور تعیین mtry و ntree



شکل ۸. اهمیت ویژگی‌های مورداستفاده در ساخت مدل جنگل تصادفی (۵ ویژگی اول) با استفاده از شاخص جینی

فشاری آسفالت توسط مدل RF برای داده‌های آموزشی و آزمایشی ترسیم شده است. با توجه به نزدیک بودن نتایج RF به خط نیم ساز، می‌توان گفت مدل مقادیر را با دقت خوبی پیش‌بینی کرده است.

پس از ساخت مدل با استفاده از بسته random forest در نرم‌افزار R نسخه ۴-۲-۱ (Breiman et al., 2015)، مدل توسط داده‌های آزمایشی مورد ارزیابی قرار گرفت که نتایج در جدول ۳ ارائه شده است. در شکل ۹ نمودار مقادیر واقعی اندازه‌گیری شده و مقادیر پیش‌بینی شده مقاومت



شکل ۹. منحنی پراکندگی مقاومت فشاری پیش‌بینی شده در مقابل مقاومت فشاری واقعی در مراحل آموزش (الف) و آزمایش (ب) در روش RF

در جدول ۳ پارامترهای بهینه برای هر مدل و نتایج ارزیابی آنها ارائه شده است. با توجه به نتایج R^2 ، SDE، RMSE و MAPE هر دو الگوریتم عملکرد موفقی در را در پیش‌بینی مقاومت فشاری آسفالت داشته‌اند که روش SVM با دارا بودن R^2 معادل ۰/۸۷۵، SDE، RMSE و MAPE معادل ۰/۰۲۰۸، ۰/۰۴۱۲ و ۳/۰۱۷۳ برای داده‌های آزمایشی نتیجه قابل قبول‌تری داشته است.

در یک مخلوط آسفالتی اگر طرح اختلاط به‌درستی کار شده باشد، به‌منظور کنترل مقاومت مارشال آسفالت، باید نسبت به تأمین قیر بهینه، درصد شکستگی و دانه‌بندی مناسب و درصد فیلر بهینه اقدام شود. براساس مطالعات صورت گرفته پارامترهای اصلی مؤثر در مقاومت مارشال، درصد قیر، درصد شکستگی و درصد فیلر می‌باشند (Aljassar et al., 2004; Ogundipe, 2016; Rashwan, 2015). با توجه به پارامترهای مؤثر در مدل‌سازی جنگل تصادفی و مطابقت آن با واقعیت، می‌توان گفت که مدل جنگل تصادفی به‌درستی توانسته روابط پیچیده بین داده‌ها را تشخیص داده و مقاومت مارشال را به‌دقت خوبی پیش‌بینی کند.

جدول ۴. نتایج ارزیابی مدل‌های RF و SVM

مدل یادگیری ماشین	پارامترها	مقدار پارامترها	R^2	SDE	RMSE	MAPE
SVM	c	۹	۰/۸۷۵	۰/۰۲۰۸	۴۰/۰۴۱۲	۳/۰۱۷۳
	γ	۰/۱۱				
Random forest	Number of trees	۱۰۰۰	۰/۸۲۶۹	۰/۰۲۱۱	۴۱/۸۷۰	۳/۱۶۴۱
	mtry	۵				

روسازی تأثیر می‌گذارد. این مطالعه کاربرد موفقیت‌آمیز یادگیری ماشین را برای پیش‌بینی مقاومت مارشال آسفالت با استفاده از آزمایش‌ها را نشان می‌دهد. دو مدل جنگل تصادفی و ماشین بردار پشتیبان برای پیش‌بینی مقاومت مارشال ساخته شد. از داده‌های آزمایشگاه کنترل کیفیت سازمان عمران شهرداری مشهد برای آموزش و ارزیابی الگوریتم‌های ML مورد استفاده قرار گرفت. معیارهای مورد استفاده دانه‌بندی مصالح، درصد شکستگی، درصد جذب قیر، وزن مخصوص قیر، وزن مخصوص حقیقی مصالح، درصد قیر مصرفی و نسبت فیلر به قیر مؤثر می‌باشند که درصد قیر، درصد شکستگی و درصد قیر (عبوری از الک ۲۰۰) بیشترین تأثیر را در عملکرد مدل داشتند. با توجه به معیارهای ارزیابی به‌دست‌آمده برای هر دو روش SVM و RF می‌توان گفت که این الگوریتم‌ها توانایی پیش‌بینی مناسب مقاومت فشاری آسفالت را دارند. در این بررسی، از روش R^2 ، RMSE، SDE، MAPE جهت ارزیابی مدل‌ها استفاده شد. مقادیر به‌دست‌آمده برای مدل SVM به ترتیب برابر با ۰/۸۷۵، ۰/۰۴۱۲، ۴۰/۰۲۰۸ و ۳/۱۰۷۳ و برای مدل RF برابر با ۰/۸۲۶۹، ۴۱/۸۷۰، ۰/۰۲۱۱ و ۳/۱۶۴۱ است؛ بنابراین می‌توان نتیجه گرفت که مدل تولیدشده

به‌طورکلی مدل‌های SVM در داده‌های پراکنده بهتر از درخت‌ها عمل می‌کنند اما تفسیرپذیری کمتری دارند کنند (Amershi et al., 2019; Andrew, 2001). در نحوه پیاده‌سازی مدل‌ها، جنگل تصادفی معمولاً سریع‌تر از مدل SVM عمل می‌کند (Longjun et al., 2011). با توجه به عملکرد قابل قبول هر دو مدل و مزایا و معایب آنها، استفاده از هر کدام بسته به نظر کاربر دارد. البته باید در نظر داشت که کلیه نمونه‌های به‌کاررفته در این پژوهش از یک نوع قیر (PG64-22) و یک نوع مصالح (مصالح آهکی) استفاده کرده‌اند و تحت یک طرح اختلاط تولید شده‌اند. به‌منظور ارزیابی بهتر مدل‌ها و مقایسه کارایی آنها بهتر است داده‌های متنوع‌تری شامل انواع مختلف قیر، مصالح که تحت طرح اختلاط‌های مختلفی تولید شده‌اند استفاده شود.

۶- نتیجه‌گیری

با توجه به رفتار پیچیده مواد روسازی آسفالتی در شرایط بارگذاری مختلف، ساختار روسازی و شرایط محیطی، پیش‌بینی دقیق پایداری روسازی آسفالت دشوار است. پیش‌بینی پایداری مارشال یکی از مهم‌ترین مؤلفه‌های عملکرد روسازی است. یک پیش‌بینی دقیق مستقیماً بر موفقیت کلیه عملکرد

machines for breast cancer classification. *Telkomnika*, 18(2), pp.815-821.

-ASTM D5444-15, "Standard Test Method for Mechanical Size Analysis of Extracted Aggregate", ASTM International, West Conshohocken, PA, 2015, www.astm.org.

-ASTM D5821-13, (2017), "Standard Test Method for Determining the Percentage of Fractured Particles in Coarse Aggregate", ASTM International, West Conshohocken, PA, www.astm.org.

-ASTM D4469-17, (2017), "Standard Practice for Calculating Percent Asphalt Absorption by the Aggregate in Asphalt Mixtures", ASTM International, West Conshohocken, PA, www.astm.org.

-ASTM D70 / D70M-21, (2021), "Standard Test Method for Specific Gravity and Density of Semi-Solid Asphalt Binder (Pycnometer Method)", ASTM International, West Conshohocken, PA, www.astm.org.

-ASTM C29 / C29M-17a, (2017), "Standard Test Method for Bulk Density, ("Unit Weight") and Voids in Aggregate, ASTM International, West Conshohocken, PA, www.astm.org.

-ASTM D2172 / D2172M-17e1, (2017), "Standard Test Methods for Quantitative Extraction of Asphalt Binder from Asphalt Mixtures", ASTM International, West Conshohocken, PA, www.astm.org.

-ASTM D6927-15, (2015), "Standard Test Method for Marshall Stability and Flow of Asphalt Mixtures", ASTM International, West Conshohocken, PA, www.astm.org

-ASTM D3515-01, (2017), "Standard Specification for Hot-Mixed", Hot-Laid Bituminous Paving Mixtures.

-Breiman, L., Cutler, A., Liaw, A., & Wiener, M., (2015), "Package random forest version 4, pp.6-12.

-Burkov, A., (2020), "Machine Learning Engineering", True Positive Inc.

-Cansiz, O. F., & Askar, D. D., (2018), "Developing Multi Linear Regression Models for Estimation of Marshall Stability", *International Journal of Advanced Engineering Research and Science*, 5, 6.

-Cardoso-Fernandes, J., Teodoro, A., Lima, A., & Roda-Robles, E., (2019), "Evaluating the performance of support vector machines (SVMs) and random forest (RF) in Li-pegmatite mapping: Preliminary results", *Earth Resources and Environmental Remote Sensing/GIS Applications X*.

با استفاده از روش SVM برای پیش‌بینی مقاومت فشاری مارشال برای داده‌های موردبررسی قابل‌اعتمادتر است. به‌منظور بهینه کردن الگوریتم‌های یادگیری ماشین به‌دست‌آمده و طراحی یک مدل پویای پیش‌بینی پیشنهاد می‌شود پس از انجام آزمایش‌ها هر نمونه جدید، از نتایج حاصل در آموزش مجدد الگوریتم استفاده شود. همچنین می‌توان از این روش‌ها به‌منظور اجتناب از آزمایش‌های وقت‌گیر و مکرر آزمایشگاهی بهره برد. این مطالعه بینشی در مورد توسعه روش‌های نوین برای بهینه‌سازی طراحی مخلوط آسفالتی برای عملکرد مخلوط مطلوب ارائه می‌کند. هنگامی که داده دارای پیچیدگی روابط هستند این روش‌ها می‌توانند این روابط پیچیده را به‌خوبی کشف کنند. با توجه به نتایج مطلوب این مطالعه، پیشنهاد می‌شود از این الگوریتم‌ها برای پیوند عملکرد روسازی با اطلاعات طراحی مخلوط، پارامترهای حجمی، ساخت و نگهداری روسازی و شناسایی شیوه‌های ساخت روسازی‌های اقتصادی و بادوام استفاده کرد. با این‌حال، باید توجه داشت که تکنیک‌های هوش مصنوعی مستقیماً از داده‌های اندازه‌گیری شده توسعه‌یافته و اعتبار داده‌ها مستقیماً بر موفقیت مدل تأثیر می‌گذارد.

۷. مراجع

-Aggarwal, C. C., (2017), "An introduction to outlier analysis", In *Outlier analysis*, Springer. pp. 1-34.

-Ali, N. S., (2006), "Flexural Behavior Compared with Marshall Stability of Asphalt Paving Mixtures", *Journal of Zankoy Sulaimani*, 9.

-Aljassar, A. H., Metwali, S., & Ali, M. A., (2004), "Effect of filler types on Marshall stability and retained strength of asphalt concrete", *International Journal of Pavement Engineering*, 5(1), pp.47-51.

-Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T., (2019), "Software engineering for machine learning: A case study", *IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*,

-Andrew, A. M., (2001), "An introduction to support vector machines and other kernel based learning methods", *Kybernetes*.

-Aroef, C., Rivan, Y., & Rustam, Z., (2020), "Comparing random forest and support vector

- Jakkula, V., (2006), "Tutorial on support vector machine (svm)", School of EECS, Washington State University, 37 .
- Jayalakshmi, T., & Santhakumaran, A. (2011), "Statistical normalization and back propagation for classification", *International Journal of Computer Theory and Engineering*, 3(1), pp.1793-8201.
- Jordan, M. I., & Mitchell, T. M., (2015), "Machine learning: Trends, perspectives, and prospects", *Science*, 349(6245), pp.255-260.
- Kanevski, M., (2009), "Machine learning for spatial environmental data: theory, applications, and software", EPFL press .
- Karahancer, S., Capali, B., Eriskin, E., Morova, N., Serin, S., Saltan, M., Terzi, S & . Kucukcapraz, D. O., (2016), "Marshall stability estimating using artificial neural network with polyparaphenylene terephthalamide fibre rate", *International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*.
- Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L., (2016), "Machine learning in geosciences and remote sensing", *Geoscience Frontiers*, 7(1), pp.3-10.
- Liaw, A., & Wiener, M., (2002), "Classification and regression by random Forest. R news, 2(3), pp.18-22.
- Liu, H., & Motoda, H., (2012), "Feature selection for knowledge discovery and data mining, Vol. 454, Springer Science & Business Media .
- Liu, X., (1998), "Strategies for outlier analysis", *IEE Colloquium on Knowledge Discovery and Data Mining*, Digest No. 1998/310.
- Longjun, D., Xibing, L., Ming, X., & Qiyue, L., (2011), "Comparisons of random forest and support vector machine for predicting blasting vibration characteristic parameters", *Procedia Engineering*, 26, pp.1772-1781.
- Majidifard, H., Adu-Gyamfi, Y., & Buttlar, W. G., (2020), "Deep machine learning approach to develop a new asphalt pavement condition index", *Construction and building materials*, 247, 118513.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C. C., Lin, C. C., & Meyer, M. D., (2019), "Package 'e1071', *The R Journal*.
- Mitchell, T. M., (1997), "Machine Learning, Burr Ridge, IL:McGraw Hill .
- Mohamad, I. B., & Usman, D., (2013), "Standardization and its effects on K-means
- Chicco, D., Warrens, M. J., & Jurman, G., (2021), "The coefficient of determination R-squared is more informative than SMAPE", MAE, MAPE, MSE and RMSE in regression analysis evaluation, *PeerJ Computer Science*, 7, e623.
- Choubane, B., Page, G. C., & Musselman, J. A., (2000), "Suitability of asphalt pavement analyzer for predicting pavement rutting", *Transportation Research Record*, 1723(1), pp.107-115.
- Chourasiya, S., & Jain, S., (2019), "A Study Review On Supervised Machine Learning Algorithms", *SSRG International Journal of Computer Science and Engineering (SSRG-IJCSE)*, 6 .
- Cortes, C & .Vapnik, V., (1995), "Support vector networks. *Machine learning*, 20(3), pp.273-297.
- Devore, J. L., (2011), "Probability and Statistics for Engineering and the Sciences", Cengage learning.
- Filzmoser, P., Garrett, R. G., & Reimann, C., (2005), "Multivariate outlier detection in exploration geochemistry", *Computers & geosciences*, 31(5), pp.579-587.
- Gareth, J., Daniela, W., Trevor, H., & Robert, T., (2013), "An introduction to statistical learning: with applications in R. Spinger.
- Ghorbanzadeh, O., Rostamzadeh, H., Blaschke, T., Gholaminia, K., & Aryal, J., (2018), "A new GIS-based data mining technique using an adaptive neuro-fuzzy inference system (ANFIS) and k-fold cross-validation approach for land subsidence susceptibility mapping", *Natural Hazards*, 94(2), pp.497-517.
- Goodfellow, I., Bengio, Y., & Courville, A., (2016), "Machine learning basics", *Deep learning*, 1(7), pp.98-164.
- Han, H., Guo, X., & Yu, H., (2016), "Variable selection using mean decrease accuracy and mean decrease gini based on random forest", 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS).
- Harris, J., & Grunsky, E. C., (2015), "Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data", *Computers & geosciences*, 80, pp.9-25 .
- Hawkins, D., (1980), "A single outlier in normal samples", In *Identification of Outliers* Springer, pp. 27-41.

- Shabankareh, M., & Hezarkhani, A., (2017), "Application of support vector machines for copper potential mapping in Kerman region", *Iran. Journal of African Earth Sciences*, 128, pp.116-126.
- Shawe-Taylor, J., & Cristianini, N., (2000), "An introduction to support vector machines and other kernel-based learning methods, Vol. 204.
- Tiwari, S., Babbar, R., & Kaur, G., (2018), "Performance evaluation of two ANFIS models for predicting water quality Index of River Satluj (India)", *Advances in Civil Engineering*.
- Wang, H., Ma, C., & Zhou, L., (2009), "A brief review of machine learning and its application, 2009 international conference on information engineering and computer science,
- Wellmer, F. W., (2012), "Statistical evaluations in exploration for mineral deposits", *Springer Science & Business Media*.
- Yang, Q., Li, X., & Shi, X., (2008), "Cellular automata for simulating land use changes based on support vector machines", *Computers & geosciences*, 34(6), pp.592-602.
- Zhang, R., Zhou, M., Gong, X., He, X., Qian, W., Qin, S., & Zhou, A., (2015), "Detecting anomaly in data streams by fractal model", *World Wide Web*, 18(5), pp.1419-1441.
- Zhang, X. D., (2020), "Machine learning, In *A Matrix Algebra Approach to Artificial Intelligence*", Springer, pp. 223-440.
- clustering algorithm", *Research Journal of Applied Sciences, Engineering and Technology*, 6(17), pp.3299-3303.
- Mohri, M., Rostamizadeh, A., & Talwalkar, A., (2018), "Foundations of Machine Learning. Adaptive Computation and Machine Learning series", In: MIT Press.
- Ogundipe, O. M., (2016), "Marshall stability and flow of lime-modified asphalt concrete. *Transportation Research Procedia*", 14, pp.685-693.
- Oshiro, T. M., Perez, P. S., & Baranauskas, J. A., (2012), "How many trees in a random forest? International workshop on machine learning and data mining in pattern recognition".
- Ozgan, E., (2009), "Fuzzy logic and statistical-based modelling of the Marshall Stability of asphalt concrete under varying temperatures and exposure times", *Advances in engineering software*, 40(7), pp.527-534.
- Özgan, E., & Saruhan, H., (2010), "Modeling of asphalt concrete via simulated annealing", *Advances in Engineering Software*, 41(4), pp.680-683.
- Rashwan, N. K., (2015), "Evaluation of aggregate and mineral filler types on characteristics of hot mix aspha", *International Journal of Civil Engineering and Technology*, 6.
- Refaeilzadeh, P., Tang, L., & Liu, H., (2009), "Cross-validation, *Encyclopedia of database systems*", 5, pp.532-538.

Predicting Marshall Asphalt Stability Using Supervised Machine Learning Algorithms, Support Vector Machine and Random Forest

Hassan Hosseinzadeh, M.Sc., Grad., Civil Engineering, Eqbal Lahori IHE, Mashhad, Iran.

Alireza Hasani, M.Sc., Student, Civil Engineering, Eqbal Lahori IHE, Mashhad, Iran.

Sina Arman, M.Sc., Grad., Civil Engineering, Islamic Azad University, Mashhad, Iran.

Amir Siavash Hejazi, M.Sc., Grad., Civil Engineering, Islamic Azad University, Mashhad, Iran.

E-mail: hosseinzadeh-ha@mashhad.ir

Received: March 2023- Accepted: August 2023

ABSTRACT

Road construction and maintenance organizations usually use certain criteria to qualify asphalt mixtures before use in construction. One of the most important features that is measured in the asphalt mixing and quality control plan is the Marshall asphalt stability. This study examines the use of machine learning techniques to predict Marshall Asphalt stability. Due to the time-consuming and costly process of asphalt production and quality control, it is necessary to use new methods in this process. In this research, two supervised support vector machine and random forest algorithms, which are machine learning algorithms, were used to predict the marshal asphalt stability. For this purpose, the test results of 2000 asphalt samples including Granulation of aggregate, Fracture percentage, bitumen adsorption, bitumen specific gravity, actual specific gravity of materials, bitumen consumption percentage, dust to effective binder ratio and Marshall asphalt stability for training and evaluation Models were used. After modeling and evaluation, the value of R2 is 87.5 for the support vector machine method and 82.69 for the random forest. Also, MAPE, RMES and SDE values for SVM were 3.1073, 40.042 and 0.0208, respectively, and for RF were 3.1641, 41.870 and 0.0211, respectively. The results show the efficiency of the models used against laboratory methods for predicting marshal asphalt stability, which SVM method has a better performance than RF. Machine learning methods can be used to predict other parameters of the asphalt mixing plan and reduce the time, cost and human error of the tests.

Keywords: Asphalt, Machine Learning, Random Forest, Support Vector Machine, Marshal Stability